

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-293317

(43)Date of publication of application : 20.10.2000

(51)Int.Cl.

G06F 3/06

G06F 12/16

G11B 19/02

G11B 20/12

(21)Application number : 2000-066061

(71)Applicant : HITACHI LTD

(22)Date of filing : 08.04.1996

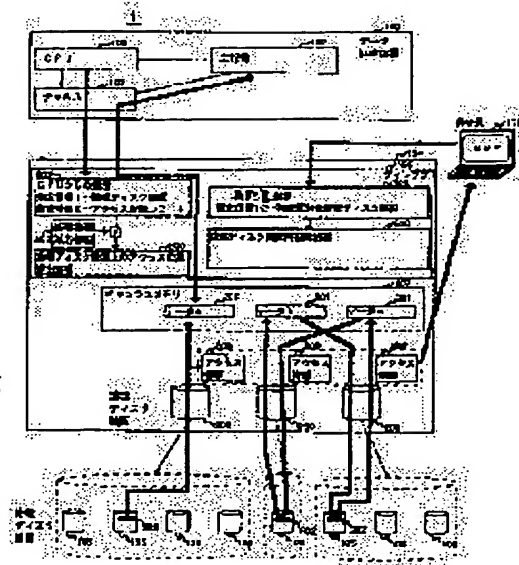
(72)Inventor : YAMAMOTO YASUTOMO
YAMAMOTO AKIRA
SATO TAKAO

(54) STORAGE CONTROLLER

(57)Abstract:

PROBLEM TO BE SOLVED: To improve the access performance in sequential access, etc., by performing rearrangement to a physical storage device in units of logical storage drives and successively storing data on the physical storage device.

SOLUTION: A service engineer refers to access information 500 presented by an SVP 111 to examine the rearrangement of the logical disk drives 200. Consequently, when there is a logical disk 200 decided to be rearranged, a rearrangement indication 620 is sent to the storage controller 104. A director 106 receives the indication 620 and performs a logical disk rearranging process 630 between two specified logical disk drives 200. At this time, logical-physical correspondence information 300 is used to transfer data in one-process units to be rearranged from the physical disk drive 105 to a cache memory 107. Then the data 201 in the process units on the cache memory 107 are repeatedly written to the physical disk drive 105 as a rearrangement destination and the correspondence information 300 is updated after the writing is completed.



LEGAL STATUS

[Date of request for examination]

27.03.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

BEST AVAILABLE COPY

[Date of registration]

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

JPO and NCPI are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The memory control unit characterized by to have a logical-memory equipment relocation means store data in the physical store of a relocation place continuously while a data processor rearranges said logical-memory equipment to said physical store based on the index defined beforehand in the memory control unit which arranges the logical memory equipment which performs direct access to the physical store which actually memorizes data, and controls the data transfer between said data processors and said physical stores.

[Claim 2] In the memory control unit with which a data processor controls the data transfer between matching, said data processor, and said physical store for the physical store which actually remembers data to be logical memory equipment which performs direct access. An access information extraction means to extract the access information to the logical memory equipment of a data processor as an index during employment of control of said data transfer. The memory control unit characterized by having a logical memory equipment relocation means to store data in the physical store of a relocation place continuously while rearranging said logical memory equipment to said physical store based on said index.

[Claim 3] The memory control unit with which said access information is characterized by including the access frequency information from said data processor to said logical memory equipment in a memory control unit according to claim 2.

[Claim 4] The memory control unit with which said access information is characterized by including the access pattern information from said data processor to said logical memory equipment in a memory control unit according to claim 2 or 3.

[Claim 5] The memory control unit characterized by said index being the dependability for which said logical memory equipment is asked in a memory control unit according to claim 1.

[Claim 6] The memory control unit characterized by providing an index presentation means to show said index to either of claim 1 to claims 5 in the memory control unit of a publication at a customer engineer, and a relocation directions reception means to receive the relocation directions from a customer engineer.

[Claim 7] The memory control unit characterized by providing a relocation directions reception means to receive the relocation directions from a data processor in a memory control unit given in either of claim 1 to claims 5.

[Claim 8] The memory control unit characterized by providing a relocation necessity decision means to determine the necessity of relocation as either of claim 1 to claims 5 in the memory control unit of a publication based on said index.

[Claim 9] When the logical memory equipment under relocation has access from a data processor in the memory control unit of a publication at either of claim 1 to claims 8, The completion field of relocation and relocation incomplete field of logical memory equipment under relocation are identified. The memory control unit characterized by providing further the access location change means which will be made to access the logical memory equipment of a relocation place if said access location is said completion field of relocation, and will be made to access the logical memory equipment concerned if said access location is said relocation incomplete field.

[Translation done.]

[Claim 10] It is the store system characterized by for said memory control unit to acquire the access situation by said data processor in the store system which has the memory control unit which controls the data transfer between two or more physical stores for holding the data of the logical store which a data processor recognizes for a data access, and said two or more physical units and data processor, and to move the data of said logical store to the 2nd physical store from the 1st physical store based on said access situation.

* NOTICES *

JPO and NCIP1 are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the memory control unit which can improve the dependability of the memory control unit which can improve the access engine performance with the case of a sequential access, or a random accelerator even when a hit ratio is low, and data in more detail about a memory control unit. Especially this invention is useful to the store subsystem constituted by the highly efficient disk unit, its highly efficient disk unit, and disk controller of the disk array sense, and the information processing system constituted by the store subsystem and data processor.

[0002]

[Description of the Prior Art] The paper "D. Patterson, G. Gibson, and R.H.Kartz:A Case for Redundant Arrays of Inexpensive Disks (RAID), ACM SIGMOD Conference, Chicago and IL, (June 1988) pp.109-116" announced in the "ACM SIGMOD" meeting held in University of Illinois of Chicago is indicating the technique about the data arrangement on a disk array.

[0003] Moreover, in JP 7-84732A, the technique of using a part of disk unit like a disk cache is indicated. It specifically divides into the temporary field which stores data for a disk unit temporarily, and the field which finally writes in data, and duplex writing of the updating data is once carried out to a temporary field, without generating parity, they carry out parity generation asynchronous, and are written in the last field.

[0004] On the other hand, the technique of changing the RAID level holding data dynamically by the difference in access frequency is indicated by the electric information American Communications Association technical research report "DE 95-68 (others [Motegi]: the performance evaluation at the time of disk failure of the disk array using Hot Mirroring, December, 1995, electric information American Communications Association technical report Vol.95-No.407, pp.19-24). By dividing a disk unit into the part of RAID1 configuration, and the part of RAID5 configuration, and specifically changing the storing location of data dynamically so that data with light access may be preferentially stored in the part of RAID1 configuration, the data with high access frequency are stored in the part of RAID1 configuration, and what has low access frequency can be stored in the part of RAID5 configuration. According to this technique, it is possible to make the physical disk equipment with which the physical disk equipment with which memory capacity differs from RAID level intermingled within a store subsystem, and the data in logical disk equipment can be stored in the physical disk equipment of arbitration based on indexes, such as that access frequency, access pattern, etc. Moreover, a storing location can also be dynamically changed so that data with high access frequency may be stored in more nearly high-speed physical disk equipment. In addition, the disk array of RAID1 writes the duplicate in the subdisk unit called a mirror to the write-in data from a data processor, and secures the dependability of data. Since redundancy data are the duplicate of the original data, the overhead of redundancy data origination is small and the access engine performance is good. However, the utilization ratio of physical storage is as low as 50%. On the other hand, the disk array of RAID5 creates the redundancy data called parity to two or more write-in data from a data processor. Since the data before updating and the parity before updating need to be led,

the overhead of redundancy data origination is large to parity creation time, and the access engine performance is bad to it. However, in order to create one parity to two or more data, the utilization ratio of storage is high compared with RAID1.

[0005]

[Problem(s) to be Solved by the Invention] With the above-mentioned conventional technique, in order to change the storing location of data by the data unit to access, on the logical disk equipment with which a data processor performs direct access, data [****] will be discontinuous on the physical disk equipment which actually memorizes data. For this reason, in the case of the sequential access which carries out read/write of a series of data, in fact, two or more data are gathered and there is a trouble which stops being able to carry out read/write and causes access performance degradation.

[0006] Many of data which the access pattern moved with the random accelerator at the part of RAID1 configuration when a hit ratio was low in order to write light data in the part of RAID which moved data judged that access frequency is low on the other hand at every light with conventional technique of the above-mentioned report "DE 95-68" to part of RAID5 from part of RAID1 configuration configuration, and was vacant 1 configuration will be again returned to the part of RAID5 configuration. For this reason, when a hit ratio is low, the improvement in the access engine performance cannot be expected, but has the trouble that the overhead of processing which moves data conversely causes access performance degradation.

[0007] Moreover, with the above-mentioned conventional technique, there is a trouble which is not taken into consideration at all about improvement in the dependability of data.

[0008] Then, with the case of a sequential access, or a random accelerator, the 1st purpose of this invention is to offer the memory control unit which can improve the access engine performance, even when a hit ratio is low. Moreover, the 2nd purpose of this invention is to offer the memory control unit which can improve the dependability of data.

[0009]

[Means for Solving the Problem] In the 1st viewpoint, a data processor arranges this invention to the physical store which actually memorizes data for the logical memory equipment which performs direct access. In the memory control unit which controls the data transfer between said data processors and said physical stores While rearranging said logical memory equipment to said physical store based on the index defined beforehand, the memory control unit characterized by having a logical memory equipment relocation means to store data in the physical store of a relocation place continuously is offered. In the memory control unit by the 1st viewpoint of the above, the storing location of data is not changed by the data unit to access, relocation to physical storage is performed for logical memory equipment as a unit, and data are continuously stored in the physical storage of a relocation place. Therefore, also in the case of a sequential access, the access engine performance can be improved. Moreover, the storing location of data is not changed at every light, but since said relocation is performed based on the index defined beforehand, even when a hit ratio is low, the access engine performance can be improved with a random accelerator.

[0010] In the 2nd viewpoint, this invention matches the physical store which actually remembers data to be logical memory equipment with which a data processor performs direct access. In the memory control unit which controls the data transfer between said data processors and said physical stores An access information extraction means to extract the access information to the logical memory equipment of a data processor as an index during employment of control of said data transfer, While rearranging said logical memory equipment to said physical store based on said index, the memory control unit characterized by having a logical memory equipment relocation means to store data in the physical store of a relocation place continuously is offered. In the memory control unit by the 2nd viewpoint of the above, the storing location of data is not changed by the data unit to access, relocation to physical storage is performed for logical memory equipment as a unit, and data are continuously stored in the physical storage of a relocation place. Therefore, also in the case of a sequential access, the access engine performance can be improved. Moreover, since the storing location of data is not changed at every light, but access information is extracted and said relocation is performed, using it

statistically, even when a hit ratio is low, the access engine performance can be improved with a random accelerator.

[0011] In the 3rd viewpoint, this invention offers the memory control unit characterized by said access information including the access frequency information from said data processor to said logical memory equipment in the memory control unit of the above-mentioned configuration. In the memory control unit by the 3rd viewpoint of the above, logical memory equipment with high access frequency is rearrangeable to more nearly high-speed physical storage. Therefore, the access engine performance can be improved.

[0012] In the 4th viewpoint, this invention offers the memory control unit characterized by said access information including the access pattern information from said data processor to said logical memory equipment in the memory control unit of the above-mentioned configuration. In the memory control unit by the 4th viewpoint of the above, logical memory equipment with the high ratio of a sequential access is rearrangeable to physical storage with the more high sequential access engine performance. Therefore, the access engine performance can be improved.

[0013] In the 5th viewpoint, this invention offers the memory control unit characterized by said index being the dependability for which said logical memory equipment is asked in the memory control unit of the above-mentioned configuration. In the memory control unit by the 5th viewpoint of the above, that it is reliable can rearrange the logical memory equipment called for to more reliable physical storage. Therefore, the dependability of data can be improved.

[0014] In the 6th viewpoint, this invention offers the memory control unit characterized by providing an index presentation means to show a customer engineer said index, and a relocation directions reception means to receive the relocation directions from a customer engineer in the memory control unit of the above-mentioned configuration. In the memory control unit by the 6th viewpoint of the above, since a customer engineer can input relocation directions, said relocation can be performed very flexibly.

[0015] In the 7th viewpoint, this invention offers the memory control unit characterized by providing a relocation directions reception means to receive the relocation directions from a data processor in the memory control unit of the above-mentioned configuration. In the memory control unit by the 7th viewpoint of the above, since a data processor can input relocation directions, a customer engineer can perform said relocation under the advanced conditions which cannot be judged.

[0016] In the 8th viewpoint, this invention offers the memory control unit characterized by providing a relocation necessity decision means to determine the necessity of relocation based on said index in the memory control unit of the above-mentioned configuration. In the memory control unit by the 8th viewpoint of the above, in order for a memory control unit to make a self-decision of the relocation directions, it is not necessary to apply a burden to a customer engineer or a data processor.

[0017] When this invention has access from a data processor in the logical memory equipment under relocation in the memory control unit of the above-mentioned configuration in the 9th viewpoint, The completion field of relocation and relocation incomplete field of logical memory equipment under relocation are identified. If said access location is said completion field of relocation, the logical memory equipment of a relocation place will be made to access, and if said access location is said relocation incomplete field, the memory control unit characterized by providing the access location change means made to access the logical memory equipment concerned will be offered. In the memory control unit by the 9th viewpoint of the above, since the completion field of relocation and relocation incomplete field of logical memory equipment under relocation are identified and the access location from a data processor is changed, while employing the data transfer between a data processor and a physical store, it is rearrangeable.

[0018] Two or more physical stores for this invention to hold the data of the logical store which a data processor recognizes for a data access in the 10th viewpoint. In the storage system which has the memory control unit which controls the data transfer between said two or more physical units and data processors said memory control unit. The access situation by said data processor is acquired, and the storage process defined system characterized by moving the data

of said logical store to the 2nd physical store from the 1st physical store based on said access situation is offered. In the memory control unit by the 10th viewpoint of the above, since the data of logical memory equipment are moved to the 1st to 2nd physical store according to the access situation by the data processor, the access engine performance can be improved.

[0019]

[Embodiment of the Invention] Hereafter, the operation gestalt of this invention is explained. In addition, thereby, this invention is not limited.

[0020] - 1st operation gestalt - The 1st operation gestalt extracts the access information of each logical disk equipment with a memory control unit, and shows it to a customer engineer through SVP (service processor), and the relocation directions of a customer engineer based on this access information perform relocation to the physical disk equipment of logical disk equipment.

[0021] Drawing 1 is the block diagram of the information processing system containing the memory control unit concerning the 1st operation gestalt of this invention. This information processing system 1 has come to connect a data processor 100, a memory control unit 104, one or more physical disk equipments 105, and SVP111.

[0022] Said data processor 100 has CPU101, the primary storage 102, and the channel 103.

[0023] Said memory control unit 104 has one or more directors 106, cache memory 107, directory 108, nonvolatile memory 109, the nonvolatile memory management information 110, the information 300 corresponding to logic physics, the logical disk equipment information 400, and access information 500. Said director 106 performs data transfer between the channel 103 of a data processor 100, the data transfer between physical disk equipment 105 and the channel 103 of a data processor 100, and said cache memory 107, and data transfer between said cache memory 107 and physical disk equipment 105. Data with the high access frequency in physical disk equipment 105 are loaded to said cache memory 107. Said director 106 performs this load processing. The examples of the data to load are the data for access of CPU101 of a data processor 100, this data for access, data with the near storing location on physical disk equipment 105, etc. Said directory 108 stores the management information of said cache memory 107. Said nonvolatile memory 109 loads data with the high access frequency in physical disk equipment 105 like said cache memory 107. Said nonvolatile memory management information 110 stores the management information of said nonvolatile memory 109. Said information 300 corresponding to logic physics is information which shows the logical disk equipment (200 of drawing 2) arranged at the location and each physical disk equipment 105 on the physical disk equipment 105 with which each logical disk equipment (200 of drawing 2) is arranged. Calculation of the storing field on the physical disk equipment 105 of the data for access of CPU101 of a data processor 100 etc. is performed using this information. Said logical disk equipment information 400 shows conditions, such as access propriety of each logical disk equipment (200 of drawing 2). Said access information 500 is the information on the access frequency of each logical disk equipment (200 of drawing 2), an access pattern, etc.

[0024] The information 300 and the logical disk information 400 corresponding to logic physics are recorded on a non-volatilized medium, in order to prevent disappearance by power off etc.

[0025] Said physical disk equipment 105 consists of a medium which records data, and equipment which write the recorded data.

[0026] Said SVP111 receives an input of the presentation to the customer engineer of access information 500 and the relocation directions 620 from a customer engineer. Moreover, presentation to customer engineers, such as a fault condition of the dispatch of directions and information processing system 1 to the information processing system 1 from a customer engineer, is performed.

[0027] Drawing 2 is drawing showing the relation of logical disk equipment 200 and physical disk equipment 105. Logical disk equipment 200 is an apparent disk unit in which CPU101 of a data processor 100 carries out direct access, and corresponds with the physical disk equipment 105 with which the data for access are actually stored. The data on logical disk equipment 200 are continuously arranged on physical disk equipment 105 in consideration of the sequential access. When the physical disk equipment 105 with which the data of logical disk equipment 200 are

arranged is a disk array configuration, this logical disk equipment 200 corresponds with two or more physical disk equipments 105. Moreover, the capacity of physical disk equipment 105 is larger than logical disk equipment 200, and when the data of two or more logical disk equipments can be stored in one physical disk equipment 105, this physical disk equipment 105 corresponds with two or more logical disk equipments 200. Correspondence of this logical disk equipment 200 and physical disk equipment 105 is managed for said information 300 corresponding to logic physics. For example, when CPU101 of a data processor 100 leads the data 201 of logical disk equipment 200, based on the information 300 corresponding to logic physics, it asks for the physical disk equipment 105 corresponding to logical disk equipment 200 with a memory control unit 104, and asks for the data storage location 202 in the field of the physical disk equipment 105, and data transfer is performed.

[0026] Drawing 3 is drawing showing the information 300 corresponding to logic physics. The information 300 corresponding to logic physics consists of logical disk configuration information 310 and physical disk configuration information 320. Said logical disk configuration information 310 is the information about the field on the physical disk equipment 105 with which each logical disk equipment 200 is arranged, and when asking for the physical disk equipment 105 which corresponds from logical disk equipment 200, it is used. On the other hand, said physical disk configuration information 320 is the information about the logical disk equipment 200 arranged at each physical disk equipment 105, and when asking for the logical disk equipment 200 which corresponds from physical disk equipment 105, it is used.

[0029] As for said logical disk configuration information 310, only the number of logical disk equipment 200 has the group of the physical disk device group 311, the RAID configuration 312, and a starting position 313. Said physical disk device group 311 is information which shows the physical disk equipment 105 with which the logical disk equipment 200 concerned is arranged. Said RAID configuration 312 shows said physical disk device group's 311 RAID level. Said starting position 313 shows the head location where the logical disk equipment 200 concerned is arranged on physical disk equipment 105.

[0030] As for said physical disk configuration information 320, only the number of physical disk equipment 105 has the logical disk device group 321. Said logical disk device group 321 shows the logical disk equipment 200 arranged at the physical disk equipment 105 concerned.

[0031] Drawing 4 is drawing showing the logical disk information 400. As for the logical disk information 400, only the number of logical disk equipment 200 has the logical disk condition 401 and the completion pointer 402 of relocation. Said logical disk condition 401 expresses the condition of the logical disk equipments 200 "normal", "lock out", "under", etc. ["under a format and relocation"] Said completion pointer 402 of relocation is effective information only while said logical disk condition 401 "is rearranging", and the head location of the field which has not yet finished relocation processing is shown, the next location 200, i.e., logical disk equipment concerned, of the field which has completed relocation processing of the logical disk equipment 200 concerned. In access to the field before the completion pointer 402 of relocation, it must access to the physical disk equipment 105 after relocation at the time of the data access "under relocation." On the other hand, in access to the field after completion pointer of relocation 402, it must access to the physical disk equipment 105 before relocation.

[0032] Drawing 5 expresses access information 500. As for access information 500, only the number of logical disk equipment 200 has the access frequency information 501 and the access pattern information 502. Refer to this access information 500 for a memory control unit 104, a data processor 100, or SVP111. Said access frequency information 501 manages the count of access to the logical disk equipment 200 concerned per unit time amount. This access frequency information 501 is used as an index which asks for the high thing or the low thing of access frequency in each logical disk equipment 200. Said access pattern information 502 manages the rate of the sequential access to the logical disk equipment 200 concerned, and random access. This access pattern information 502 has many sequential accesses, and rearranging to physical disk equipment 105 with the more high sequential engine performance uses it as an index which asks for desirable logical disk equipment 200.

[0033] Next, actuation of a memory control unit 104 is explained. Drawing 6 is drawing which

expressed actuation of a memory control unit 104 to the detail. First, the actuation at the time of read/write processing is explained. In case a director 106 usually performs read/write processing, he receives the directions 600 from CPU from CPU101 via a channel 103. The directions 600 from this CPU include the assignment information 1 which specifies the logical disk equipment 200 with which the record for a lead (or light) is memorized, and the assignment information 2 which specifies the location (a truck, a sector, record) in the logical disk equipment 200 with which the record for a lead (or light) is memorized. A director 106 is access location calculation processing (610) on physical disk equipment, and computes the access location on physical disk equipment 105 using the directions 600 from said CPU, and the information 300 corresponding to logic physics. This physical disk drive-access location calculation processing (610) is explained in full detail with reference to drawing 8 later. Then, for example by lead processing, the data of the data storage location 202 on the computed physical disk equipment 105 are read out on cache memory 107, it considers as data 201, and the read-out data 201 is transmitted to a primary storage 102 through a channel 103.

[0034] Next, extraction processing of access information 500 is explained. At the time of access of the read/write processing from CPU101, a director 106 updates the access information 500 of the logical disk equipment 200 for access. Extraction of the access frequency information 501 counts up the internal counter at every access, and judges access frequency from said internal UNTA at the time of access after access progress of fixed time amount or the count of fixed. Extraction of the access pattern information 502 counts up the count of a sequential access at the internal counter at every access, and judges an access pattern from said internal counter at the time of access after access progress of fixed time amount or the count of fixed.

[0035] Next, the relocation directions 620 are explained. A customer engineer examines the need for relocation of each logical disk equipment 200 with reference to the access information 500 shown through SVP111. If there is logical disk equipment 200 which opted for relocation as a result of this examination, the relocation directions 620 consist of directions information 1-2 which 104 through SVP111. These relocation directions 620 consist of directions information 1-2 which specifies two logical disk equipments 200 for relocation. The contents of examination which a customer engineer performs are the same as that of logical disk equipment relocation necessity decision processing (910) in which it explains with reference to drawing 10 with the 3rd operation gestalt mentioned later.

[0036] Next, logical disk equipment relocation processing (630) is explained. A director 106 performs logical disk equipment relocation processing (630) between two specified logical disk equipments 200 in response to said relocation directions 620. Drawing 7 is the processing flow Fig. of the logical disk equipment relocation processing section 630. At step 700, the logical disk condition 401 of two logical disk equipments 200 of having been specified of the logical disk information 400 is set up "during relocation." At step 701, the completion pointer 402 of relocation of two logical disk equipments 200 with which it was specified of the logical disk information 400 is initialized in the head location of each logical disk equipment 200. At step 702, the completion pointer 402 of relocation of two logical disk equipments 200 with which it was specified of the logical disk information 400 is checked, and if relocation of all fields is not completed, and it progressed to step 703 and has completed, it will progress to step 707.

[0037] At step 703, data transfer from physical disk equipment 105 to a cache memory 107 top is performed to the data for 1 time of the batch of relocation processing from the data location which the completion pointer 402 of relocation shows. Here, the amount of data for 1 time of a batch is determined as the least common multiple of each amount of data corresponding to one redundancy data of two logical disk equipments 200 for relocation. For example, if it rearranges between the logical disk equipment 200 of RAID5, and the logical disk equipment 200 of RAID1, since the amount of data corresponding to one redundancy data of the logical disk equipment 200 of RAID1 is "1", the amount of data for 1 time of a batch will be determined as the amount of data corresponding to one redundancy data of the logical disk equipment 200 of RAID5, i.e., the amount of data corresponding to one parity.

[0038] At step 704, when the relocation place logical disk equipment 200 of each logical disk equipment 200 for relocation is the thing of RAID level which has parity, parity is generated to

the data 201 for 1 time of the batch for [on cache memory 107] relocation. At step 705, the parity created at the data 201 and said step 704 for 1 time of a batch for relocation is written in the physical disk equipment 105 of a relocation place. [on cache memory 107] At step 706, the completion pointer 402 of relocation is carried forward by 1 time of the batch. And it returns to said step 702.

[0039] In addition, in the above-mentioned step 703, 704, it transmits and doubles also to nonvolatile memory 109, and data and parity prevent data missing by the cache failure. This reason at the time of the writing in the above-mentioned step 705 For example, the inside of the data of the 1st logical disk equipment 200 and the 2nd logical disk equipment 200. Supposing the data on cache memory 107 become access impossible according to a failure in the phase which wrote the data of the 1st logical disk equipment 200 in physical disk equipment 105 (physical disk equipment 105 with which the dimension is arranged at the 2nd logical disk equipment 200) It is because the data of the 2nd logical disk equipment 200 which writing has not ended disappear (the data of the 1st logical disk equipment 200 will be overwritten as mentioned above by the physical disk equipment 105 with which the dimension is arranged at the 2nd logical disk equipment 200).

[0040] At step 707, the information 300 corresponding to logic physics is updated. That is, the logical disk configuration information 310 and the physical disk configuration information 321 are changed. At step 708, the logical disk condition 401 of the logical disk information 400 is returned to the original condition, and relocation processing (630) is ended.

[0041] Next, physical disk drive-access location calculation processing (610) is explained. Drawing 8 is the processing flow Fig. of the physical disk drive-access location calculation processing section 610. If it confirms whether the logical disk condition 401 of the logical disk equipment 200 for access of the logical disk information 400 "is rearranging" and becomes "during relocation" about it at step 800 -- step 801 -- progressing -- "under relocation -- it is --" -- if there is nothing, it will progress to step 803.

[0042] The completion pointer 402 of relocation and the access data location of the logical disk equipment 200 for access of the logical disk information 400 are compared, if an access data location becomes after the location which the completion pointer 402 of relocation points out, it will progress to step 802, and if an access data location becomes a front [location / which the completion pointer 402 of relocation points out], it will progress to step 803 at step 801.

[0043] At step 802, the logical disk equipment 200 of the relocation place of the logical disk equipment 200 concerned is made applicable to access. And it progresses to step 804.

[0044] At step 803, the logical disk equipment 200 concerned is made applicable to access.

[0045] At step 804, the access location on the physical disk equipment 105 corresponding to the logical disk equipment 200 for access is computed using the information 300 corresponding to logic physics.

[0046] According to the information processing system 1 and the memory control unit 104 concerning the above operation gestalt [1st], logical disk equipment with high access frequency is rearrangeable to more nearly high-speed physical disk equipment with the decision of a customer engineer based on access information 500. Moreover, logical disk equipment with the high ratio of a sequential access is rearrangeable to physical disk equipment with the more high sequential access engine performance. Therefore, the access engine performance can be improved.

[0047] -- The operation gestalt of the 2nd operation gestalt--above 1st is transformed, and access information 500 is shown to a data processor 100 from a memory control unit 104, and a data processor 100 determines relocation necessity and you may make it take out relocation directions (about [620]) to a memory control unit 104.

[0048] -- 3rd operation gestalt -- The 3rd operation gestalt receives relocation directions neither from SVP111 nor a data processor 100, but a memory control unit 104 makes a self-decision.

[0049] Drawing 9 is drawing which expressed actuation of a memory control unit 104 to the detail. The difference from the 1st operation gestalt (drawing 6) is that the logical disk relocation necessity decision processing section 910 issues the relocation directions 620.

[0050] Drawing 10 is the processing flow Fig. of the above-mentioned logical disk relocation

necessity decision processing section 910. A director 106 performs this logical disk relocation necessity decision processing (910) by inspecting the access information 500 of each logical disk equipment 200 a fixed period. If there is [whether there is logical disk equipment / low speed / the physical disk equipment 105 which access frequency exceeds default value and is arranged at step 1000 with reference to the access frequency information 501 on access information 500 / equipment / comparatively / (this is hereafter called 1st candidate logical disk equipment) 200, and] logical disk equipment 200 which checks and corresponds, it will progress to step 1001, and if there is nothing, it will progress to step 1005.

[0051] At step 1001, it confirms whether the ratio of a sequential access is beyond default value, with reference to the access pattern information 502 on said 1st candidate logical disk equipment 200, if it is not beyond default value, it will progress to step 1002, and with default value [beyond], it progresses to step 1004.

[0052] At step 1002, with reference to the access frequency information 501 on the logical disk equipment 200 arranged at physical disk equipment 105 more nearly high-speed than said 1st candidate logical disk equipment 200, if check and it is [whether access frequency has logical disk equipment 200 below default value (this is hereafter called 2nd candidate logical disk equipment), and], it will progress to step 1003, and if there is nothing, it will progress to step 1005.

[0053] It determines that relocation processing (630) is required between said 1st candidate logical disk equipment 200 and said 2nd candidate logical disk equipment 200, and the relocation directions 620 are taken out with step 1003. And processing is ended.

[0054] At step 1004, with reference to the access pattern information 502 on the logical disk equipment 200 arranged at physical disk equipment 105 with the sequential engine performance higher than said 1st candidate logical disk equipment 200, if check and it is [whether the ratio of a sequential access has logical disk equipment 200 below default value (this is hereafter called 2nd candidate logical disk equipment), and], it will progress to said step 1003, and if there is nothing, it will progress to said step 1002.

[0055] At step 1005, it is determined that relocation processing (630) of logical disk equipment 200 is unnecessary. And processing is ended.

[0056] According to the information processing system 1 and the memory control unit 104 concerning the above operation gestalt [3rd], based on access information 500, logical disk equipment with high access frequency is rearrangeable to more nearly high-speed physical disk equipment automatically. Moreover, logical disk equipment with the high ratio of a sequential access is rearrangeable to physical disk equipment with the more high sequential access engine performance. Therefore, the access engine performance can be improved.

[0057] -- 4th operation gestalt--above-mentioned the 1-- the 3rd operation gestalt -- deforming -- access information 500 -- replacing with -- or -- in addition, the dependability required of logical disk equipment 200 may be used for the index of relocation processing necessity decision. If dependability is used for an index, the dependability of the data on logical disk equipment 200 can be raised.

[0058]

[Effect of the Invention] According to the memory control unit of this invention, with the case of a sequential access, or a random accelerator, even when a hit ratio is low, the access engine performance can be improved. Moreover, according to the memory control unit of this invention, the dependability of data can be improved.

[Translation done.]

* NOTICES *

JP0 and NCPI are not responsible for any damages caused by the use of this translation.

- 1. This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

- [Drawing 1] It is the block diagram of the information processing system containing the memory control unit concerning the 1st operation gestalt of this invention.
- [Drawing 2] It is the explanatory view of the correspondence relation between logical disk equipment and physical disk equipment.
- [Drawing 3] It is the configuration instantiation Fig. of the information corresponding to logic physics.
- [Drawing 4] It is the configuration instantiation Fig. of logical disk information.
- [Drawing 5] It is the configuration instantiation Fig. of access information.
- [Drawing 6] It is the block diagram showing actuation of the memory control unit in the 1st operation gestalt of this invention.
- [Drawing 7] It is the processing flow Fig. of the logical disk equipment relocation processing section.
- [Drawing 8] It is the processing flow Fig. of the physical disk drive-access location calculation processing section.
- [Drawing 9] It is the block diagram showing actuation of the memory control unit in the 3rd operation gestalt of this invention.
- [Drawing 10] It is the processing flow Fig. of the logical disk equipment relocation necessity decision processing section.
- [Description of Notations]
- 1 --- Information Processing System
- 100 --- Data processor
- 101 --- CPU
- 102 --- Primary storage
- 103 --- Channel
- 104 --- Memory control unit
- 105 --- Physical disk equipment
- 106 --- Director
- 107 --- Cache memory
- 108 --- Cache directory
- 109 --- Nonvolatile memory
- 110 --- Nonvolatile memory management information
- 111 --- SVP
- 200 --- Logical disk equipment
- 201 --- Data
- 202 --- Data storage location
- 300 --- Information corresponding to logic physics
- 400 --- Logical disk information
- 500 --- Access information
- 600 --- Directions from CPU
- 610 --- The access location calculation processing section on physical disk equipment

- 620 --- Directions information
- 630 --- Logical disk equipment relocation processing section
- 910 --- Logical disk relocation necessity decision processing section

[Translation done.]

(19)日本特許庁 (J P) (12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-293317
(P2000-293317A)

(43)公開日 平成12年10月20日(2000.10.20)

識別記号	G 0 6 F 3 / 0 6	P I	チーフ・ド(参考)	
			G 0 6 F 3 / 0 6	3 0 2 J
G 0 6 F 3 / 0 6	3 0 2		3 0 1 S	
	5 4 0		5 4 0	
	1 2 / 1 6		3 2 0 L	
G 1 1 B 1 9 / 0 2	3 2 0		5 0 1 K	
G 1 1 B 1 9 / 0 2	5 0 1			

審査請求 未請求 請求項の数10 OL (全 14 頁) 最終頁に続く

(21)出願番号	特開2000-66061(P2000-66061)	(71)出願人	000005108
(62)分割の表示	特開平8-853700の分割	株式会社日立製作所	
(22)出願日	平成8年4月8日(1996.4.8)	東京都千代田区神田豊町4丁目6番地	
		山本 威友	
		特来川川崎市麻生区王禅寺1099番地 株	
		式会社日立製作所システム開発研究所内	
		山本 彰	
		特来川川崎市麻生区王禅寺1099番地 株	
		式会社日立製作所システム開発研究所内	
		佐藤 孝夫	
		神奈川県小田原市国府津2880番地 株式会	
		社日立製作所ストレージシステム事業部内	
		井理士 有正 神志郎	
		100095511	

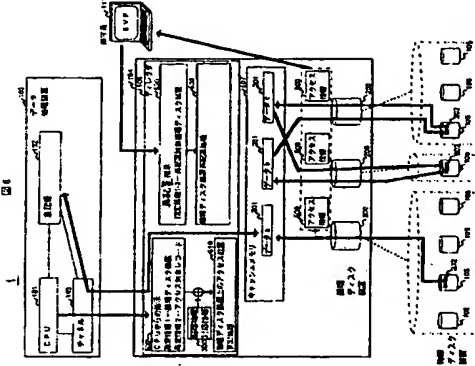
(54) [発明の名称] 記憶制御装置

(57) [要約]

【課題】 アクセス性能を向上する。

【解決手段】 各論理ディスク装置200に対するアクセス情報500を採取し、そのアクセス情報500をS V P 1 1 を通じて保守員に提示する。保守員の再配置指示620があると、指示された2つの論理ディスク装置200の間で物理ディスク装置105を配置替えし、全データを連想的に格納し直す。

【効果】 アクセス頻度の高い論理ディスク装置をより高速な物理ディスク装置へ再配置することが出来る。シーク・ジャンプ・アクセスの比率の高い論理ディスク装置をよりシーク・ジャンプ・アクセス性能の高い物理ディスク装置へ再配置することが出来る。



【特許請求の範囲】

【請求項1】 データ処理装置が直接アクセスを行う論理的記憶装置を複数にデータを記憶する物理的記憶装置に配置し、前記データ処理装置と前記物理的記憶装置とのデータ転送を制御する記憶制御装置において、予め定められた指図に基づいて前記論理的記憶装置を物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを連想的に格納する論理的記憶装置再配置手段を有することを特徴とする記憶制御装置。

【請求項2】 データ処理装置が直接アクセスを行う論理的記憶装置と装置にデータを記憶する物理的記憶装置とを対応付け、前記データ処理装置と前記物理的記憶装置との間のデータ転送を制御する記憶制御装置において、前記データ処理装置の運用中にデータ処理装置の論理的記憶装置へのアクセス情報を指図として採取するアクセス情報採取手段と、前記指図に基づいて前記論理的記憶装置を前記物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを連想的に格納する論理的記憶装置再配置手段とを有することを特徴とする記憶制御装置。

【請求項3】 請求項2に記載の記憶制御装置において、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶制御装置。

【請求項4】 請求項3に記載の記憶制御装置において、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶制御装置。

【請求項5】 請求項1に記載の記憶制御装置において、前記指図が、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶制御装置。

【請求項6】 請求項1から請求項5のいずれかに記載の記憶制御装置において、前記指図を保守員に提示する指図提示手段と、保守員からの再配置指示を受け付ける再配置指示受付手段とを具備したことを特徴とする記憶制御装置。

【請求項7】 請求項1から請求項5のいずれかに記載の記憶制御装置において、データ処理装置からの再配置指示を受け付ける再配置指示受付手段を具備したことを特徴とする記憶制御装置。

【請求項8】 請求項1から請求項5のいずれかに記載の記憶制御装置において、前記指図に基づいて再配置の要否を決定する再配置要否決定手段を具備したことを特徴とする記憶制御装置。

【請求項9】 請求項1から請求項8のいずれかに記載の記憶制御装置において、再配置中の論理的記憶装置にデータ処理装置からのアクセスがあったとき、再配置中の論理的記憶装置の再配置完了領域と再配置未完了領域とを識別し、前記アクセス位置を前記再配置完了領域ならは再配置先の論理的記憶装置にアクセスさせ、前記ア

セス位置が前記再配置未完了領域ならば当該論理的記憶装置にアクセスさせるアクセス位置切替手段をさらに具備したことを特徴とする記憶制御装置。

【請求項10】 データ処理装置がデータアクセスのために記憶する論理的記憶装置のデータを保持するため複数の物理的記憶装置と、前記複数の物理的記憶装置とデータ処理装置との間のデータ転送を制御する記憶制御装置とを有する記憶制御システムにおいて、前記記憶制御装置は、前記データ処理装置によるアクセス状態を取得し、前記アクセス状態に基づいて前記論理的記憶装置のデータを第1の物理的記憶装置から第2の物理的記憶装置に移動させることを特徴とする記憶制御システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、記憶制御装置に関する。さらに詳しくは、シーク・ジャンプ・アクセスの場合やランダムアクセスでヒット率が低い場合でもアクセス性能を向上することが出来る記憶制御装置およびデータの信頼性を向上することが出来る記憶制御装置に関する。

特に、本発明は、ディスクレイアウト向きの高機能ディスク装置、その高機能ディスク装置とディスク制御装置とにより構成される記憶制御サブシステム、およびその記憶制御サブシステムとデータ処理装置とにより構成される情報処理システムに有用である。

【0002】

【従来の技術】 シカゴのイリノイ大学で開かれた「ACM SIGMOD」会議において発表された論文「D. Patterson, G. Gibson, and R. H. Kartz, A Case for Redundant Arrays of Inexpensive Disks (RAID) ACM SIGMOD Conference, Chicago, IL, June 1988, pp. 109-116」は、ディスクレイ上のデータ配置に関する技術を開示している。

【0003】 また、特開平7-84732号公報では、ディスク装置の一部をディスクキャッシュの如く用いる技術が開示されている。具体的には、ディスク装置を一時的にデータを格納するテンポラリ領域と最終的にデータを書き込む領域とに分け、更新データはバリティを生成せずに一旦テンポラリ領域に二重書きし、非同期的にバリティ生成し、最終領域に書き込む。

【0004】 一方、電気情報通信学会技術研究報告「D E 9 5 - 6 8 (茂木他: Hot Mirroring を用いたディスクアレのディスク故障時の性能評価, 1995年12月、電気情報通信学会技術報告 Vol. 95-No. 407, pp. 19-24)」には、アクセス頻度の高いデータは、データを保持するRAIDレベルを動的に変更する技術が開示されている。具体的には、ディスク装置をRAID1構成の部分とRAID5構成の部分に分け、ライトアクセスのあったデータを優先的にRAID1構成の部分に格納するようにデータの格納位置を動的に変更することにより、アクセス頻度の高いデータはRAID1構成の部分に格

納し、アクセス頻度の低いものはRAID5構成の部分に格納するよう出来る。この技術によれば、記憶容量の質なる物理ディスク装置やRAIDレベルの質なる物理ディスク装置を記憶装置サブシステム内で混在させることが可能であり、論理ディスク装置内のデータを、そのアクセス頻度やアクセスパターンなどの指標に基づいて、任意の物理ディスク装置に格納することが出来る。また、アクセス頻度の高いデータを、より高速な物理ディスク装置に格納するよう、動的に格納位置を変更することも出来る。なお、RAID1のディスクアレイは、データ処理装置からの書き込みデータに対して、その複製をミラーと呼ばれる別ディスク装置に書き込み、データの信頼性を確保する。冗長データ作成のヘッドが小さく、アクセス性能が良い。但し、物理的記憶装置の使用率は、50%と低い。一方、RAID5のディスクアレイは、データ処理装置からの複製の書き込みデータに対して、パーティと呼ばれる冗長データを作成する。パーティ作成時に更新前データと更新前パーティのリードに必要なため、冗長データ作成のオーバーヘッドが大きく、アクセス性能は悪い。但し、複製のデータに対して一つのパーティを作成するため、記憶装置の使用効率はRAID1に比べ高い。

【0005】
【発明が解決しようとする課題】上記従来の技術では、アクセスするデータ単位でデータの格納位置の変更を行うため、データ処理装置が直接アクセスを行う論理ディスク装置上では遅様なデータが、実際のデータを配置する物理ディスク装置上では非連続となってしまう。このため、一のデータをリード/ライトするランケンジャンルアクセスの場合、実際には複数データまでまとめてリード/ライトできなくなり、アクセス性能の低下を招く問題点がある。

【0006】一方、上記報告「DE 95-68」の従来技術では、ライトの度に、アクセス頻度が低いと判断して、データを RAID1 構成の部分から RAID5 構成の部分に移し、空いた RAID1 構成部分をランダムアクセスで書き込むため、アクセスパターンがランダムな場合でヒット率が低い場合には、RAID1 構成の部分に移したデータの多くは再び RAID5 構成の部分に移れることになる。このため、ヒット率が低い場合、アクセス性能の向上は期待できず、逆にデータを移す処理のオーバーヘッドがアクセス性能の低下を引き起こす問題点がある。

【0007】また、上記の従来技術では、データの信頼性の向上については全く考慮されていない問題点がある。

【0008】そこで、本発明の第1の目的は、シートケン
シャルアクセスの場合やランダムアクセスでヒット率が低い場合でも、アクセシビリティを向上することが出来る。

制御装置を提供することにある。また、本発明の第2の目的は、データの信頼性を向上することが出来る記憶制御装置を提供することにある。

【0009】
 課題を解決するための手段】第1の観点では、本発明は、データ処理装置に直接アクセスを行う論理的記憶装置を、データ処理装置と物理的記憶装置との間のデータ転送を制御する記憶制御装置において、予め定めた指標に基づいて前記論理的記憶装置を前記物理的記憶装置に再配置すると共に再配置後の物理的記憶装置にデータを逆移格納する論理的記憶装置を再配置手段を有することを特徴とする記憶制御装置を提供する。上記第1の観点とを特徴とする記憶制御装置では、アクセスするデータ単位でデータの格納位置の変更を行うのではなく、論理的記憶装置を単位として物理的記憶装置への再配置を行い、且つ、再配置先の物理的記憶装置にデータを逆移格納しても、アクセス性能を向上することが出来る。また、ライトの度にデータの格納位置の変更を行うのではなく、予め定めた指標に基づいて前記再配置を行うから、ランダムアクセス性能を向上させることが出来る。

【0010】第2の観点では、本発明は、データ処理装置が直接アクセスを行う論理的記憶装置と実際にデータ処理装置と物理的記憶装置とを対応付け、前記データ処理装置と物理的記憶装置の間のデータ転送を制御する記憶制御装置において、前記データ転送の制御の適用中にデータ処理装置の論理的記憶装置へのアクセス情報報を指標として採取するアクセス情報報取得手段と、前記指標に基づいて前記記憶装置を前記物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを再読取に格納する論理的記憶装置を記憶手段とを有することと特徴とする記憶制御装置を提供する。上記第2の観点によする記憶制御装置では、アクセスするデータ単位であるデータの格納位置の変更を行うのではなく、論理的記憶装置を単位として物理的記憶装置への再配置を行い、且つ、再配置先の物理的記憶装置にデータを逆読取に格納する。従って、シークンシャルアクセスの場合でも、アクセス性能を向上することが出来る。また、ライトの度にデータの格納位置の変更を行うのではなく、アクセス情報報を採取し、それを統計的に利用して前記再配置を行ってから、ランダムアクセスで読み取りの場合でも、アクセス性能を向上することが出来る。

【0011】第3の観点では、本発明は、上記構成の記憶制御装置において、前記アクセス情報が、前記データ処理装置から前記論理の記憶装置へのアクセス頻度情報を含むことと特徴とする記憶制御装置を提供する。

第3の観点による記憶制御装置では、アクセス頻度の高い論理の記憶装置をより高度な物理的記憶装置へ再配置し、

することが出来る。従って、アクセス性能を向上すること
とが出来る。

【0012】第4の観点では、本発明は、上記構成の配
置制御装置から前記アクセス情報が、前記データ
情報を含むことを特徴とする記述制御装置を提供す
る。上記第4の観点による記述制御装置では、シークンシャ
ルアクセスの比率の高い論理的記憶装置をよりシークン
シャルアクセス性の高い物理的記憶装置へ再配置するこ
とができる。従って、アクセス性能を向上させることが
出来る。

【0013】第5の観点では、本発明は、上記構成の記憶制御装置において、前記指標が、前記論理的記憶装置に求められる信頼性であることと特徴とする記憶制御装置を提供する。上記第5の観点による記憶制御装置では、信頼性が高いことが求められる論理的記憶装置による信頼性の低い論理的記憶装置へ再配置することが出来る。従って、データの信頼性を向上することが出来る。

【0014】第6の観点では、本発明は、上記構成の記憶制御装置において、前記指標を保守員に提示する指標提示手段と、保守員からの再配置指示を受け付ける再配置指示受付手段とを具備したことを特徴とする記憶制御装置を提供する。上記第6の観点による記憶制御装置では、保守員が再配置指示を入力できるため、非常に柔軟に前記再配置を行うことが出来る。

【0015】第7の観点では、本発明は、上記構成の記憶制御装置において、データ処理装置からの再配置指示を受け付ける再配置指示受付手段を具備したことを特徴とする記憶制御装置を提供する。上記第7の観点によ力、記憶制御装置では、データ処理装置が再配置指示を入力できるため、保守員では判断不可能な高度の条件下で再配置を行うことが出来る。

【0016】第8の観点では、本発明は、上記構成の配役制装置において、前記指図に基づいて再配置の要否を特定するとともに、前記指図に基づいて再配置の要否を決定する再配置要否決定手段を具備したことを特徴とする配役制装置を提供する。上記第8の観点による配役制装置では、配役制装置が再配置指示を自己決定して実行するため、保守員やオペレータ処理装置に負担をかけることなく、

【0017】第9の観点では、本発明は、上記構成の制御装置において、再配置中の論理的記装装置にデータ処理装置からのアクセスがあったとき、再配置中の論理的記装装置の再配置完了領域と再配置未了領域とを識別し、前記アクセス位置が前記再配置完了領域ならば再配置前の論理的記装装置にアクセスさせ、前記アクセス位置が前記再配置未了領域ならば当該論理的記装装置はアクセスさせずアクセス位置切替手段を具備したことを特徴とする記装制御装置を提供する。上記第9の観点によれば、再配置中の論理的記装装置の再配置完了領域と再配置未了領域とを識別し、再配置完了領域と再配置未了領域とを識別し、

装置からのアクセス位置を切り替えるから、データ処理装置と物理的記憶装置の間のデータ転送を運用中に再配置を行うことが出来る。

【0018】第100の観点からは、本発明は、データ処理装置がデータアクセスのために記憶する論理的な記憶装置のデータと、前記記憶装置とデータ処理装置との間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、前記記憶制御装置は、前記データ処理装置によるアクセス状況に基づき、前記アクセス状況に基づいて前記論理的な記憶装置のデータを第1の物理的な記憶装置から第2の物理的な記憶装置に移動させることと特徴とする記憶制御装置を、上記第100の観点による記憶制御装置では、データ処理装置によるアクセス状況に応じた論理的記憶装置のデータを第1から第2の物理的な記憶装置へと移動させるから、アクセス性を向上することになる。

【6100】

【発明の実施の形態】以下、本発明の実施形態を説明する。なお、これにより本発明が限定されるものではない。

【0020】 第1の実施形態—
第1の実施形態は、各論理ディस्क装置のアクセス情報
を記憶制御装置で探知し、SVP（サービスプロセッ
サ）を通じて保守員に提示し、このアクセス情報に基づ
く保守員の再配置指示により、論理ディस्क装置の物理
ディस्क装置への再配置を行うものである。

【0021】図1は、本発明の第1の実施形態にかかる記憶制御装置を含む情報処理システムのブロック図である。この情報処理システム1は、データ処理装置100と、記憶制御装置104と、1台以上の物理ディスク装置105と、SVPI11とを接続してなっている。

【0022】前記データ処理装置100は、CPU101と、主記憶102と、チャネル103とを有している。

【0023】前記配線制御装置104は、つ以上のデータレクタ106と、揮発性メモリモジュール107と、不揮発性メモリモジュール109と、データ処理装置105と、物理情報110と、論理物理対応情報300と、論理ディスク装置情報400と、アクセス情報500を有格成している。前記データレクタ106は、データ処理装置100のデータレクタ103と物理ディスク装置105の間のデータ転送、データ処理装置100のデータレクタ103と前記キャッシュメモリモジュール107との間のデータ転送を行う。前記キャッシュメモリモジュール105中のアクセス頻度の高いデータをロードしておく。このロード処理は、前記データレクタ106が実行する。ロードするデータのデータレクタ106は、データ処理装置100のCPU1010がアクセスする

録データや、このアクセス対象データと物理ディस्क装置105上の格納位置に近いデータ等である。前記ディスク装置108は、前記キャッシュメモリ107の管理情報格納部109に、前記不揮発性メモリ109は、前記キャッシュメモリ107と同様に、物理ディスク装置105の中のアクセス頻度の高いデータをロードしておく。前記不揮発性メモリ管理情報110は、前記不揮発性メモリ109の管理情報を格納する。前記論理物理対応情報300は、各管理ディスク装置(図2の200)が配置されている物理ディスク装置105上の位置および各物理ディスク装置105に配置されている論理ディスク装置(図2の200)を示す情報である。この情報を用いて、データ処理装置100のCPU101のアクセス対象データの物理ディスク装置105上の格納領域の算出などを行う。前記論理ディスク装置情報400は、各論理ディスク装置(図2の200)のアクセス可否等の状態を示す。前記アクセス情報500は、各論理ディスク装置(図2の200)のアクセス頻度やアクセスパターンなどの情報である。

【0024】論理物理対応情報300と論理ディスク情報400は、電源断などによる消失を防ぐために不揮発性の媒体に記録する。

【0025】前記物理ディスク装置105は、データを記録する媒体と、記録されたデータを読み書きする装置とから構成される。

【0026】前記SVP111は、アクセス情報500の保守員への提示や、保守員からの再配図指示620の入力の受け付けを行う。また、保守員からの情報処理システム1への指示の発行や、情報処理システム1の障害や状態等の保守員への提示を行う。

【0027】図2は、論理ディスク装置200と物理ディスク装置105の関連を表わした図である。論理ディスク装置200は、データ処理装置100のCPUユニット1が直接アクセス可能な見掛け上のディスク装置で、アクセス対象データが実際に格納される物理ディスク装置200上のデータは、シーケンシャルアクセスを考慮して、物理ディスク装置105上に系統的に配置されている。論理ディスク装置200は複数の物理ディスク105と対応する。物理ディスク105がディスクアレイ構成の場合、該論理ディスク装置200は複数の物理ディスク105と対応する。また、物理ディスク装置200より大きく、複数の論理ディスク装置200より小さく、複数の物理ディスク装置105に格納される場合、一台を1台の物理ディスク装置105と対応する場合には、該物理ディスク装置105は複数の論理ディスク装置200と対応する。この論理ディスク装置200は物理ディスク装置105の対応は前記論理ディスク装置200で管理される。例えば、データ処理装置100で300で管理される。例えば、データ処理装置100で

CPU101が論理ディスク装置200のデータ20をリードする時、配制御装置104で論理物理対応付

アクセスポスターの情報5.0.2.2を、論理ディスク装置2.0.0の数だけ有している。このアクセス情報5.0.0は、配線制装置104、データ処理装置100、SVPI 111のいずれからも参照することが出来る。前記アクセス情報5.0.1は、車庫間あたりの当該論理ディスク装置2.0.0へのアクセス回数を管理する。このアクセス情報5.0.1は、各論理ディスク装置2.0.0の中でアクセスポスターの低いもの又は低いものを求める指標として用いる。前記アクセススポスター情報5.0.2は、当該論理ディスク装置2.0.0へのシークンシャルアクセスとラダムアクセスの割合を管理する。このアクセススポスター情報5.0.2は、シークンシャルアクセスが多く、よりシークンシャル性能の高い物理ディスク装置1.0.5に再配置することが望ましい。論理ディスク装置2.0.0を求める指標として用いる。

【0033】次に、配線制御装置104の動作を説明する。図6は、配線制御装置104の動作を詳細に表した図である。まず、リード/ライト処理時の動作について説明する。ディレクタ106は、通常リード/ライト処理を実行する際、CPU101からチャネル103を經由してPPUからの指示600を受け取る。このCPUからの指示600は、リード（またはライト）対象のレコードが記憶されている論理ディスク装置200を指定する指定情報1と、リード（またはライト）対象のレコードが記憶されている論理ディスク装置200内の位置を指定する指定情報2と、リード（またはライト）対象のレコードが記憶されている論理ディスク装置200内の位置（トラック、セクタ、レコード）を指定する指定情報3とを有している。ディレクタ106は、物理ディスク装置2上のアクセス位置算出処理（610）で、前記CPUからの指示600と論理物理対応情報300とを用いて、物理ディスク装置105上のアクセス位置を算出する。この物理ディスク装置105上の位置をキャッシュ記憶105上のデータ格納位置202のデータをキャッシュメモリ107上に読み上げてデータ201とし、その読み上げたデータ201をチャネル103を通じて主記憶102に伝送する。

【0034】次に、アクセス情報500の採取処理について説明する。CPU101からのリード/ライト処理の際にアクセス時に、ディレクタ106は、アクセス対象論理ディスク装置200のアクセス情報500を更新する。アクセス速度情報501の採取は、例えば、アクセスの度に内部カウンタをカウントアップし、一定時間または一定回数のアクセス経過後のアクセス時に、前記内部カウンタからアクセス回数を読み出すことにより行う。また、バタースン情報502の採取は、例えば、アクセスの度にバタースンカウンタを増加させ、一定時間または一定回数のアクセス経過後の前記バタースンカウンタからバタースン値を読み出すことにより行う。

【0035】次に、再配置指示620を説明する。保守員は、SVPI11を通じて提示されたアクセス情報510を参照して、各論理ディスク装置200の再配置に必要な条件を検討する。この検討の結果、再配置を決定した論理ディスク装置200があれば、SVPI11を通じて配制御装置104に対して再配置指示620を出す。この再配置指示620は、再配置対象の論理ディスク装置200を2つ指定する指示情報1-2からなる。保守員が行う検討の内容は、後述する第3の実施形態で図10を参照して説明する論理ディスク装置再配置要否検査処理(910)と同様である。

【0036】次に、論理ディスク装置再配置処理（630）を説明する。ステップ106は、制記再配置指示000を受けて、指定された2つの論理ディスク装置200の間で論理ディスク装置再配置処理630を行う。図7は、論理ディスク装置再配置処理630の処理フロー図である。ステップ700では、論理ディスク情報400のうちの指定された2つの論理ディスク装置200の論理ディスク状態401を「再配置中」に設定する。ステップ701では、論理ディスク情報400のうちの指定された2つの論理ディスク装置200の先頭位置に初期化する。ステップ702では、論理ディスク情報400のうちの指定された2つの論理ディスク装置200の再配置完了ポイント402をチェックし、全領域の再配置完了していない場合はステップ703へ進む。否、完了している場合はステップ707へ進む。

【0037】ステップ703では、再配置完了ポイン
402が示すデータ位置から再配置処理の1回の処理単
位のデータに対して物理ディスク装置105からキャ
ッチメモリ107上へのデータ転送、転送を行う。ここで、
1回の処理単位分のデータ量は、再配置対象の2つの論
理ディスク装置200の冗長データ1つに対応する各デ
ータ量の最小公倍数に決定される。たとえば、再配置を
RAID55の論理ディスク装置200とRAID10の論
理ディスク装置200の間で行うならば、RAID10の論
理ディスク装置200の冗長データ1つに対応するデー
タ1つに対応するデータ量となる。

【0038】ステップ704では、再配対対象の名前処理ディスタス表200の再配対先処理ディスタス表200の場合、カブリティを有するRAIDレベルのみの場合、キャッシュメモリ107上の再配対表101の処理単位（再配対データ201）の再配対先処理ディスタス表200に対してカブリティを抽出する。ステップ705では、キャッシュメモリ107上の再配対表101の再配対先処理ディスタス表200および前記再配対ディスタス表101の処理単位（再配対データ201）と再配対先処理ディスタス表200で作成したカブリティを、再配対先の物理ディスク装置105へ書き込む。ステップ706では、1回の

(8)

11
処理単位分だけ再配置完了ポインタ402を進める。そして、前記ステップ702に遷る。
【0039】なお、上記ステップ703、704において、データおぼろげリテリィは、不揮発性メモリ109にも転送して二重化し、キャッシュ履歴によるデータ消失を防ぐ。この理由から、上記ステップ705での書き込み時に、例えば、第1の論理ディスク装置200と第2の論理ディスク装置200のデータのうち、第1の論理ディスク装置200のデータを物理ディスク装置105（元は第2の論理ディスク装置200に配置されていた物理ディスク装置105）へ書き込んだ段階で障害によりキャッシュメモリ107上のデータがアクセス不能になったとすると、書き込みが終了してない第2の論理ディスク装置200のデータが消失するからである（元は第2の論理ディスク装置200に配置されていた物理ディスク装置105には、上記のように第1の論理ディスク装置200のデータが上書きされている）。

10
【0040】ステップ707では、論理物理対応情報310を更新する。すなわち、論理ディスク構成情報310と物理ディスク構成情報321を変更する。ステップ708では、論理ディスク情報400の論理ディスク状態401を元の状態に戻し、再配置処理（630）を終了する。

20
【0041】次に、物理ディスク装置アクセス位置算出処理（610）の説明を図8は、物理ディスク装置アクセス位置算出処理610の処理フロー図である。ステップ800では、論理ディスク情報400のうちのアクセス対象論理ディスク装置200の論理ディスク状態401が「再配置中」であるか否かをチェックし、「再配置中」ならばステップ801に進み、「再配置中」でなければステップ803に進む。

30
【0042】ステップ801では、論理ディスク情報400のうちのアクセス対象論理ディスク装置200の再配置完了ポインタ402とアクセスデータ位置とを比較し、アクセスデータ位置が再配置完了ポインタ402の指す位置以後ならばステップ802に進み、アクセスデータ位置が再配置完了ポインタ402の指す位置より前ならばステップ803に進む。

40
【0043】ステップ802では、当該論理ディスク装置200の再配置先の論理ディスク装置200をアクセス対象にする。そして、ステップ804へ進む。
【0044】ステップ803では、当該論理ディスク装置200をアクセス対象とする。

50
【0045】ステップ804では、アクセス対象の論理ディスク装置200に対応した物理ディスク装置105上でのアクセス位置を、論理物理対応情報300を用いて算出する。
【0046】以上の第1の実施形態にかかる情報処理システム1および記憶制御装置104によれば、アクセス情報500に基づく保守員の判断により、アクセス頻度の高い論理ディスク装置をより高度な物理ディスク装置へ再配置することが出来る。また、シーケンシャルアクセスの比率の高い論理ディスク装置をよりシーケンシャルアクセス性能の高い物理ディスク装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【0047】一第2の実施形態—
上記第1の実施形態を变形して、記憶制御装置104からアクセス情報500をデータ処理装置100に提示し、データ処理装置100が再配置可否を決定し記憶制御装置104に再配置指示（620相当）を出すようにしてもよい。

【0048】一第3の実施形態—
第3の実施形態は、再配置指示をSVP111やデータ処理装置100から受けるのではなく、記憶制御装置104が自己決定するものである。
【0049】図9は、記憶制御装置104の動作を詳細に表わした図である。第1の実施形態（図6）との違いは、論理ディスク再配置可否決定処理910が再配置指示620を出すことである。

【0050】図10は、上記論理ディスク再配置可否決定処理910の処理フロー図である。この論理ディスク再配置可否決定処理（910）は、ディレクタ106が一定周期で各論理ディスク装置200のアクセス情報500を検査して行う。ステップ1000では、アクセス情報500のアクセス頻度情報501を参照し、アクセス頻度が規定値を超え且つ配置されている物理ディスク装置105が比較的低速なものである論理ディスク装置（以下、これを第1候補論理ディスク装置という）200があるか否かをチェックし、該当する論理ディスク装置200があればステップ1001へ進み、なければステップ1005へ進む。

【0051】ステップ1001では、前記第1候補論理ディスク装置200のアクセスパターン情報502を参照し、シーケンシャルアクセスの比率が規定値以上であるか否かをチェックし、規定値以上でなければステップ1002へ進み、規定値以上でなければステップ1004へ進む。

13
【0054】ステップ1004では、前記第1候補論理ディスク装置200よりシーケンシャル性能の高い物理ディスク装置105に配置されている論理ディスク装置200のアクセスパターン情報502を参照し、シーケンシャルアクセスの比率が規定値以下の論理ディスク装置（以下、これを第2候補論理ディスク装置という）200があるか否かをチェックし、あれば前記ステップ1003へ進み、なければ前記ステップ1002へ進む。

20
【0055】ステップ1005では、論理ディスク装置200の再配置処理（630）は不要であると決定する。そして、処理を終了する。

【0056】以上の第3の実施形態にかかる情報処理システム1および記憶制御装置104によれば、アクセス情報500に基づいて自動的に、アクセス頻度の高い論理ディスク装置をより高度な物理ディスク装置へ再配置することが出来る。また、シーケンシャルアクセスの比率の高い論理ディスク装置をよりシーケンシャルアクセス性能の高い物理ディスク装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【0057】一第4の実施形態—
上記第1〜第3の実施形態を变形して、アクセス情報500に代えて又は加えて、論理ディスク装置200に要求される信頼性を再配置可否決定の指標に用いてもよい。信頼性を指標に用いれば、論理ディスク装置200上のデータの信頼性を向上させることが出来る。

【0058】
【発明の効果】本発明の記憶制御装置によれば、シーケンシャルアクセスの場合やランダムアクセスでヒット率が低い場合でも、アクセス性能を向上することが出来る。また、本発明の記憶制御装置によれば、データの信頼性を向上することが出来る。

【図面の簡単な説明】
【図1】本発明の第1の実施形態にかかる記憶制御装置を含む情報処理システムのブロック図である。
【図2】論理ディスク装置と物理ディスク装置との対応関係の説明図である。
【図3】論理物理対応情報の構成例説明図である。

【図4】

論理ディスク情報

図4

図5

アクセス情報

【図5】



14
【図4】論理ディスク情報の構成例説明図である。
【図5】アクセス情報の構成例説明図である。
【図6】本発明の第1の実施形態における記憶制御装置の動作を示すブロック図である。
【図7】論理ディスク装置再配置処理部の処理フロー図である。
【図8】物理ディスク装置アクセス位置算出処理部の処理フロー図である。
【図9】本発明の第3の実施形態における記憶制御装置の動作を示すブロック図である。
【図10】論理ディスク装置再配置可否決定処理部の処理フロー図である。

【符号の説明】
1 ……情報処理システム
100…データ処理装置
101…CPU
102…主記憶
103…チャネル
104…記憶制御装置
105…物理ディスク装置
106…ディレクタ
107…キャッシュメモリ
108…キャッシュディレクトリ
109…不揮発性メモリ
110…不揮発性メモリ管理情報
111…SVP
200…論理ディスク装置
201…データ
202…データ格納位置
300…論理物理対応情報
400…論理ディスク情報
500…アクセス情報
600…CPUからの指示
610…物理ディスク装置上のアクセス位置算出処理部
620…指示情報
630…論理ディスク装置再配置処理部
910…論理ディスク再配置可否決定処理部

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.